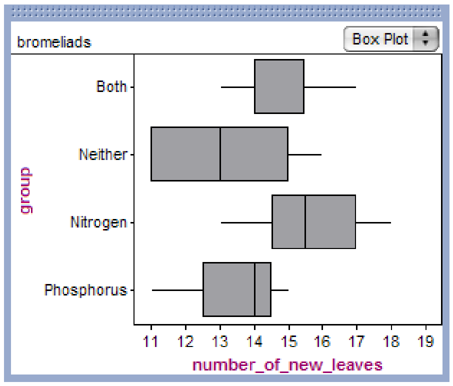
Chapter 13

Section 13.1

Check Your Understanding, Page 11:

1. Call the actual mean number of new leaves that would grow on bromeliads like the ones in this study μ_1 for the nitrogen treatment, μ_2 for the phosphorus treatment, μ_3 for the treatment using both, and μ_4 for the control treatment. The null hypothesis is $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ and the alternative hypothesis is H_a: not all of μ_1, μ_2, μ_3, and μ_4 are equal.

2. Random: The researchers randomly assigned the plants to the treatments. 10%: Not necessary because we are not sampling without replacement. Normal/Large Sample: Boxplots of the number of new leaves in each of the four treatment groups are given below. Although there is some skewness in the phosphorus, neither, and "both" groups (in the "both" group, the median is equal to the first quartile at 14), there are no outliers and the skewness is not strong enough to prevent the use of one-way ANOVA.



Same SD: The sample standard deviations for the four treatment groups are $s_1 = 1.685$, $s_2 = 1.414$, $s_3 = 1.302$, and $s_4 = 2.059$. These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{2.059}{1.302} = 1.581 < 2$. All conditions have been met so we can carry out a one-way ANOVA to compare the mean number of new leaves for the four treatments.

Check Your Understanding, Page 12:

1. The total number of observations is $N = 8 + 8 + 8 + 7 = 31$. There were 4 treatments so k = 4. The numerator degrees of freedom are $k - 1 = 4 - 1 = 3$, and the denominator degrees of freedom are $N - k = 31 - 4 = 27$.

Check Your Understanding, Page 15:

1. If the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ of no difference in mean number of new leaves is true, there is a 0.031 probability of getting differences among the sample mean number of new leaves as large as or larger than the ones observed in the experiment just by the chance involved in the random assignment.

2. Because the *P*-value of 0.031 is smaller than α = 0.05, we would reject H_0. We have convincing evidence that the four treatments do not result in the same mean number of new leaves.

Case Closed, Page 21:

1. The distribution of sugar is skewed to the right for cereals on the bottom shelf, skewed to the left for cereals on the middle shelf, and roughly symmetric for cereals on the top shelf. Typically, cereals on the middle shelf have the most sugar, followed by cereals on the top shelf and then cereals on the bottom shelf. The amount of variability is about the same in all three distributions and none have outliers.

2. H_0: $\mu_T = \mu_M = \mu_B$ versus H_a: the means are not all the same, where μ_T is the true mean amount of sugar in cereals from the top shelf, μ_M is the true mean amount of sugar in cereals from the middle shelf, and μ_B is the true mean amount of sugar in cereals from the bottom shelf.

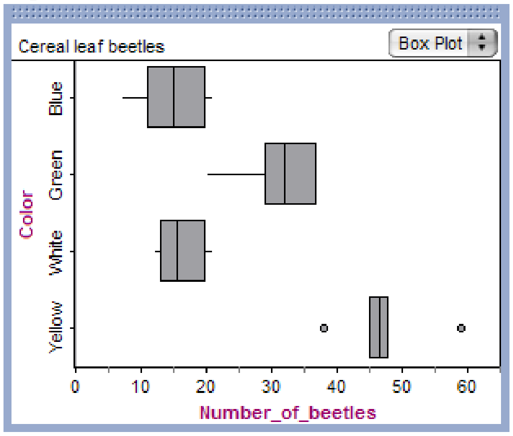
3. *Random*: Three independent random samples were selected, one from each shelf.
10%: We must assume that there are at least 10(36) = 360 cereal brands on the top shelf, 10(21) = 210 cereal brands on the middle shelf, and 10(19) = 190 cereal brands on the bottom shelf. These quantities seem a little unrealistic, so we will proceed with caution.
Normal/Large Sample: Two of the sample sizes (middle, bottom) are less than 30. However, neither of these distributions shows strong skewness or outliers, so it is reasonable to proceed.
Equal SD: The ratio of the largest SD to the smallest SD is 4.483/3.836 = 1.17 is less than 2.

4. Test statistic: F = 6.60, dfNumer = 2, dfDenom = 73, P-value = 0.002. Because the P-value of 0.002 is less than α = 0.05, we reject H_0. There is convincing evidence that the true mean amount of sugar is different for cereals on at least one shelf.

5. Because the confidence intervals don't overlap, we have convincing evidence that there is a difference in the true mean amount of sugar for cereals on the middle and bottom shelves.

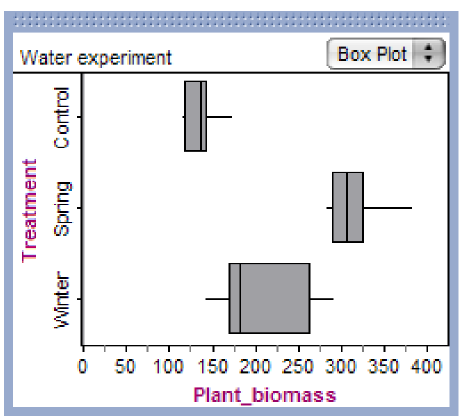
Exercises, Page 24:

13.1 (a) Call the true mean number of beetles μ_1 for blue, μ_2 for green, μ_3 for white, and μ_4 for yellow. The null hypothesis is $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ and the alternative hypothesis is H_a: not all of μ_1, μ_2, μ_3, and μ_4 are equal.
(b) Random: The researchers randomly assigned the locations of the boards of each color. 10%: Not necessary because there was no sampling without replacement. Normal/Large Sample: Boxplots of the number of beetles in each of the four treatment groups are given below. The skewness in the blue and green groups is probably small enough to not worry us. However, the outliers in the yellow group are a problem. The Normal/Large Sample condition is not met.



Same SD: The sample standard deviations for the four treatment groups are $s_1 = 5.34$, $s_2 = 6.31$, $s_3 = 3.76$, and $s_4 = 6.79$. These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{6.79}{3.76} = 1.806 < 2$. The Normal/Large Sample condition has not been met so we cannot carry out a one-way ANOVA to compare the mean number of beetles for the four treatments.

13.2 (a) Call the true mean amount of plant biomass μ_1 for winter, μ_2 for spring, and μ_3 for control. The null hypothesis is $H_0: \mu_1 = \mu_2 = \mu_3$ and the alternative hypothesis is H_a: not all of μ_1, μ_2, and μ_3 are equal.
(b) Random: The researchers randomly assigned the treatments to plots of land. 10%: Not necessary because there was no sampling without replacement. Normal/Large Sample: Boxplots of the amount of biomass in each of the three treatment groups are given below. Although there is some skewness to all three groups, there are no outliers and the skewness is not strong enough to prevent the use of one-way ANOVA.



Same SD: The sample standard deviations for the three treatment groups are $s_1 = 58.8$, $s_2 = 37.3$, and $s_3 = 21.69$. These standard deviations do not satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{58.8}{21.69} = 2.711 > 2.$ The Same SD condition was not met so we cannot carry out a one-way ANOVA to compare the mean amount of biomass for the three treatments.

13.3 Using ANOVA would not be wise in this case because there is no random selection involved. The researchers used the entire population of parents in this valley. Because all of the parents came from the same location we cannot generalize to families in other locations. Because all families in the one location were used, there is no need for inference.

13.4 Using ANOVA would not be wise in this case because there is no random selection involved. The researchers used the entire population of managers or professionals at one large manufacturing firm. Because all of the observations came from the same company we cannot generalize to men who work for other companies. Because all men in the one company were used, there is no need for inference.

13.5 (a) Call the mean score on the "angry/threatening driving scale" for the population of drivers in each of the three age groups μ_1 for less than 30, μ_2 for between 30 and 55, and μ_3 for older than 55. The null hypothesis is $H_0 : \mu_1 = \mu_2 = \mu_3$ and the alternative hypothesis is
H_a : not all of μ_1, μ_2, and μ_3 are equal.
(b) Random: Even though there was only one random sample, it is reasonable to treat these as three independent random samples. 10%: Because sampling without replacement was used, there must be more than $10(244) = 2440$ drivers under the age of 30, $10(734) = 7340$ drivers between the ages of 30 and 55, and $10(364) = 3640$ drivers over the age of 55. This seems safe to assume. Normal/Large Sample: $244 \geq 30$, $734 \geq 30$, and $364 \geq 30$. Same SD: The sample standard deviations for the four treatment groups are $s_1 = 3.11$, $s_2 = 2.21$, and $s_3 = 1.60$. These standard deviations satisfy our rule of thumb
$\frac{\text{largest } s}{\text{smallest } s} = \frac{3.11}{1.60} = 1.944 < 2.$

13.6 (a) Randomly assign 144 fruit flies to the four treatments: 1 mg/ml caffeine, 2.5 mg/ml of caffeine, 5 mg/ml of caffeine, no caffeine. There will be 36 fruit flies in each of the 4 groups. Measure the number of minutes of rest for each fruit fly during the 12-hour dark period.
(b) Call the actual mean number of minutes of rest for fruit flies like the ones in this study μ_1 for 1 mg/ml caffeine, μ_2 for 2.5 mg/ml caffeine, μ_3 for 5 mg/ml caffeine, and μ_4 for no caffeine. The null hypothesis is $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ and the alternative hypothesis is H_a : not all of μ_1, μ_2, μ_3, and μ_4 are equal.
(c) Random: The researchers randomly assigned fruit flies to treatments. Normal/Large Sample: Each of the group sizes is large ($36 \geq 30$).
(d) If the data confirm that the Same SD condition is met, that means that the ratio of the largest sample standard deviation to the smallest sample standard deviation is less than 2.

13.7 The parameters of interest are the true mean difference in memory score between 8 a.m. and 9 p.m. for morning people, evening people, and people who are neither. Call the true mean difference in memory score for the three groups μ_1 for morning people, μ_2 for evening people, and μ_3 for people who are neither. Because there are three groups of interest, $k = 3$. The sample sizes of the three groups are $n_1 = 16$, $n_2 = 30$, and $n_3 = 54$. This means that $N = 16 + 30 + 54 = 100$. The ANOVA *F*-statistic has $k - 1 = 3 - 1 = 2$ numerator degrees of freedom and $N - k = 100 - 3 = 97$ denominator degrees of freedom.

13.8 The parameters of interest are the true mean quality score for essays written by students who receive one of four treatments: those who receive no additional instruction (Group A), those who are required to prepare a written outline (Group B), those who are given 15 ideas that might be relevant to the essay topic (Group C), and those who are both given the ideas and required to prepare the outline (Group D). Call the true mean quality score for the four groups μ_1 for Group A, μ_2 for Group B, μ_3 for Group C, and μ_4 for Group D. Because there are four groups of interest, $k = 4$. The sample size for each group was $n_i = 20$. This means that $N = 20 + 20 + 20 + 20 = 80$. The ANOVA *F*-statistic has $k - 1 = 4 - 1 = 3$ numerator degrees of freedom and $N - k = 80 - 4 = 76$ denominator degrees of freedom.

13.9 The parameters of interest are the true mean math scores for children with learning disabilities who receive one of four treatments: problems read by a teacher (Group A), problems read by a computer (Group B), problems read by a computer that also shows a video (Group C), and students read the problems themselves as in an ordinary testing situation (Group D). Call the true mean math score for the four groups μ_1 for Group A, μ_2 for Group B, μ_3 for Group C, and μ_4 for Group D. Because there are four groups of interest, $k = 4$. The sample size for each group was $n_i = 25$. This means that $N = 25 + 25 + 25 + 25 = 100$. The ANOVA *F*-statistic has $k - 1 = 4 - 1 = 3$ numerator degrees of freedom and $N - k = 100 - 4 = 96$ denominator degrees of freedom.

13.10 The parameters of interest are the true mean triglyceride level for normal-weight men, overweight men, and obese men. Call the true mean triglyceride level for the three groups μ_1 for normal-weight men, μ_2 for overweight men, and μ_3 for obese men. Because there are three groups of interest, $k = 3$. The sample sizes of the three groups are $n_1 = 719$, $n_2 = 885$, and $n_3 = 220$. This means that $N = 719 + 885 + 220 = 1824$. The ANOVA *F*-statistic has $k - 1 = 3 - 1 = 2$ numerator degrees of freedom and $N - k = 1824 - 3 = 1821$ denominator degrees of freedom.

13.11 (a) If the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ of no difference in the true mean score on an "angry/threatening driving scale" is true, there is less than a 0.01 probability of getting differences among the sample mean scores as large as or larger than the ones observed in the study just by the chance involved in the random selection of drivers.
(b) Because the *P*-value of < 0.01 is smaller than α = 0.05, we reject H_0. We have convincing evidence that the true mean score on the "angry/threatening driving scale" is different for at least one of the age groups of drivers.

13.12 (a) If the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ of no difference in the true mean amount of rest during the dark period is true, there is less than a 0.0001 probability of getting differences among the sample mean scores as large as or larger than the ones observed in the study just by the chance involved in the random assignment.
(b) Because the *P*-value of less than 0.0001 is smaller than α = 0.05, we reject H_0. We have convincing evidence that the true mean amount of rest is different for at least one of the four treatments.

13.13 (a) Because there are three treatment groups (the three different makeups of the students that will be hosted), $k = 3$. There are 90 black students who are assigned to the three different treatment groups so $N = 90$. This means that the ANOVA *F*-statistic has $k - 1 = 3 - 1 = 2$ numerator degrees of freedom and $N - k = 90 - 3 = 87$ denominator degrees of freedom.
(b) Using technology, the *P*-value = Fcdf(lower:2.47, upper:10000, dfNumer:2, dfDenom:87) = 0.0905. If the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ of no difference in the true mean likelihood of playing the type of music preferred by the other race is true, there is a 0.0905 probability of getting differences among the

sample mean scores as large as or larger than the ones observed in the study just by the chance involved in the random assignment.
(c) Because the *P*-value of 0.0905 is greater than α = 0.05, we fail to reject the null hypothesis. We do not have convincing evidence that the racial mix of the gathering affects the choice of music when black students like these are planning the party.

13.14 (a) Because there are three treatment groups (the three different makeups of the students that will be hosted), $k = 3$. There are 96 white students who are assigned to the three different treatment groups so $N = 96$. This means that the ANOVA *F*-statistic has $k-1=3-1=2$ numerator degrees of freedom and $N-k=96-3=93$ denominator degrees of freedom.
(b) Using technology, the *P*-value = Fcdf(lower:16.48, upper:10000, dfNumer:2, dfDenom:93) < 0.0001. If the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3$ of no difference in the true mean likelihood of playing the type of music preferred by the other race is true, there is less than a 0.0001 probability of getting differences among the sample mean scores as large as or larger than the ones observed in the study just by the chance involved in the random assignment.
(c) Because the *P*-value of less than 0.0001 is less than α = 0.05, we reject the null hypothesis. We have convincing evidence that the racial mix of the gathering affects the choice of music when white students like these are planning the party.

13.15 (a) Random: The researchers randomly selected students of different racial/ethnic groups. 10%: Because sampling without replacement was used, there must be more than $10(809)=8090$ African American students, $10(1860)=18{,}600$ White students, $10(654)=6540$ Asian/Pacific Islander students, $10(883)=8830$ Hispanic students, and $10(207)=2070$ Native American students in the United States. This seems safe to assume. Normal/Large Sample: 809 ≥ 30, 1860 ≥ 30, 654 ≥ 30, 883 ≥ 30, 207 ≥ 30. Same SD: The sample standard deviations for the five racial/ethnic groups are $s_1 = 1.40$, $s_2 = 1.36$, $s_3 = 1.32$, $s_4 = 1.31$, and $s_5 = 1.28$. These standard deviations satisfy our rule of thumb

$$\frac{\text{largest } s}{\text{smallest } s} = \frac{1.40}{1.28} = 1.094 < 2.$$

(b) $$\bar{x} = \frac{809(2.57)+1860(2.32)+654(2.63)+883(2.51)+207(2.51)}{809+1860+654+883+207} = \frac{10{,}850.25}{4413} = 2.459.$$

$$MSG = \frac{809(2.57-2.459)^2+1860(2.32-2.459)^2+654(2.63-2.459)^2+883(2.51-2.459)^2+207(2.51-2.459)^2}{5-1}$$

$$= \frac{67.863}{4} = 16.966.$$

$$MSE = \frac{808(1.4)^2+1859(1.36)^2+653(1.32)^2+882(1.31)^2+206(1.28)^2}{4413-5} = \frac{8010.9842}{4408} = 1.817.$$

$$F = \frac{16.966}{1.817} = 9.337.$$

(c) Use an *F* distribution with 5 – 1 = 4 numerator degrees of freedom and 4413 – 5 = 4408 denominator degrees of freedom. Because the *P*-value of less than 0.001 is less than α = 0.05, we reject the null hypothesis. There is convincing evidence that the true mean interest in mathematics is different for at least one racial/ethnic group.

13.16 (a) Random: The researchers randomly assigned subjects to one of the three treatments: a single long exercise period 5 days per week, several 10-minute exercise periods 5 days per week, and several 10-minute periods 5 days per week on a home treadmill that was provided. 10%: Not necessary because there was no sampling without replacement. Normal/Large Sample: $37 \geq 30$, $36 \geq 30$, $42 \geq 30$. Same SD: The sample standard deviations for the three treatment groups are $s_1 = 4.2$, $s_2 = 4.5$, and $s_3 = 5.2$.

These standard deviations satisfy our rule of thumb $\dfrac{\text{largest } s}{\text{smallest } s} = \dfrac{5.2}{4.2} = 1.238 < 2.$

(b) $\bar{x} = \dfrac{37(10.2)+36(9.3)+42(10.2)}{37+36+42} = \dfrac{1140.6}{115} = 9.918.$

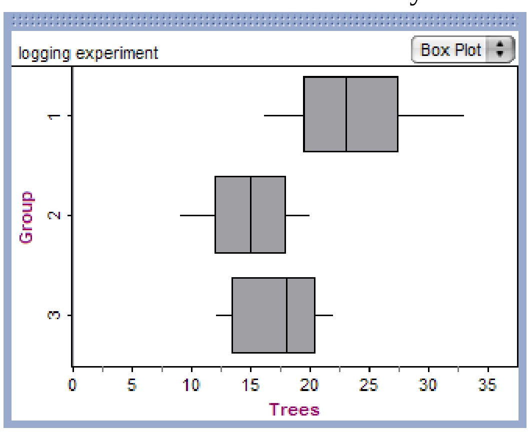
$$MSG = \frac{37(10.2-9.918)^2 + 36(9.3-9.918)^2 + 42(10.2-9.918)^2}{3-1} = \frac{20.03166}{2} = 10.0158.$$

$$MSE = \frac{36(4.2)^2 + 35(4.5)^2 + 41(5.2)^2}{115-3} = \frac{2452.43}{112} = 21.897.$$

$$F = \frac{10.0158}{21.897} = 0.457.$$

(c) Use an *F* distribution with 3 – 1 = 2 numerator degrees of freedom and 115 – 3 = 112 denominator degrees of freedom. Because the *P*-value of 0.634 is greater than α = 0.05, we fail to reject the null hypothesis. There is not convincing evidence that at least one true mean weight loss is different for subjects like these who are assigned one of the three treatments.

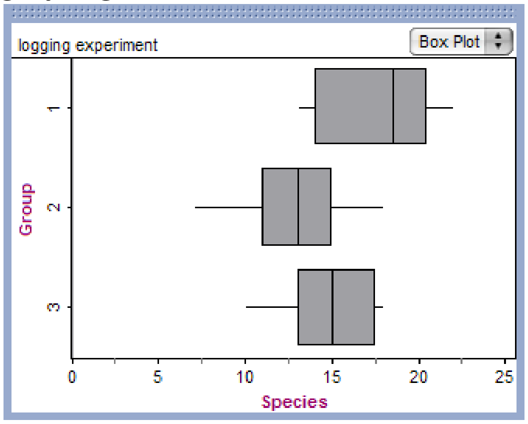
13.17 (a) The boxplots below show that the plots that have never been logged typically have more trees than the other two types of plots. The plots that were logged 8 years ago seem to have somewhat more trees than those logged just a year ago. The distributions of number of trees from all three types of plots are relatively symmetric with about the same amount of variability and have no outliers.



(b) *State:* We want to perform a test of $H_0: \mu_1 = \mu_2 = \mu_3$ versus H_a: not all of μ_1, μ_2, and μ_3 are equal, where μ_1 = the true mean number of trees in plots that had never been logged, μ_2 = the true mean number of trees in plots that had been logged 1 year earlier, and μ_3 = the true mean number of trees in plots logged 8 years earlier. *Plan:* If conditions are met, we should perform an ANOVA *F* test to compare the means. Random: The plots were randomly selected. 10%: We have data from less than 10% of all possible forest plots under these three conditions in Borneo. Normal/Large Sample: The boxplots in part (a) show no strong skewness or outliers. Same SD: The sample standard deviations for the three groups are $s_1 = 5.065$, $s_2 = 3.371$, and $s_3 = 3.955$. These standard deviations satisfy our rule of thumb $\dfrac{\text{largest } s}{\text{smallest } s} = \dfrac{5.065}{3.371} = 1.503 < 2.$ The conditions for ANOVA are met. *Do:* From the Minitab

output we have $F = 12.62$ and P-value = 0.000. *Conclude:* Because the P-value of approximately 0 is less than α = 0.05, we reject the null hypothesis. We have convincing evidence that at least one true mean number of trees is different for plots subjected to different logging treatments.

13.18 (a) The boxplots below show that the plots that have never been logged typically have more species of trees than the other two types of plots. The plots that were logged 8 years ago seem to have somewhat more species than those logged just a year ago. The distributions of number of species of trees from all three types of plots are relatively symmetric and have no outliers, although the distribution for the never logged plots has a slightly larger *IQR* than the other two.



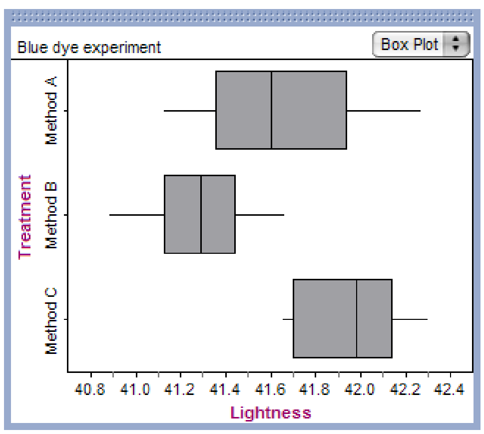
(b) *State:* We want to perform a test of $H_0: \mu_1 = \mu_2 = \mu_3$ versus H_a: not all of μ_1, μ_2, and μ_3 are equal, where $\mu_1 =$ the true mean number of tree species in plots that had never been logged, $\mu_2 =$ the true mean number of tree species in plots that had been logged 1 year earlier, and $\mu_3 =$ the true mean number of tree species in plots logged 8 years earlier. *Plan:* If conditions are met, we should perform an ANOVA F test to compare the means. Random: The plots were randomly selected. 10%: We have data from less than 10% of all possible forest plots under these three conditions in Borneo. Normal/Large Sample: The boxplots in part (a) show no strong skewness or outliers. Same SD: The sample standard deviations for the three groups are $s_1 = 3.529$, $s_2 = 3.264$, and $s_3 = 2.850$. These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{3.529}{2.850} = 1.238 < 2$. The conditions for ANOVA are met. *Do:* From the Minitab output we have $F = 6.35$ and P-value = 0.005. *Conclude:* Because the P-value of 0.005 is less than α = 0.05, we reject the null hypothesis. We have convincing evidence that at least one true mean number of species of trees is different for plots subjected to different logging treatments.

13.19 Design A allows the use of one-way ANOVA to compare the lists because the subjects are assigned to one of the 4 lists completely at random. Design B does not allow the use of one-way ANOVA to compare the lists because the treatments aren't assigned to subjects completely at random. Instead, each subject is treated as a block and listens to all four lists. *Note*: Two-way ANOVA could be used to analyze this design, but this method is beyond the scope of this chapter.

13.20 (a) Use a two-sample t test for this experiment. There are two treatments being compared with a quantitative response variable. Alternatively, a one-way ANOVA could also be used here but it has an additional condition (same SD). The two-sample t test does not require the standard deviations to be the same.
(b) Use a one-way ANOVA for this experiment. There are three treatments being compared with a quantitative response variable.
(c) Use a chi-square test for homogeneity for this experiment. There are three treatments being compared with a categorical response variable (whether or not they choose to work with someone else).

13.21 *State:* We want to perform a test of $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ versus H_a: not all of μ_1, μ_2, μ_3, and μ_4 are equal, where $\mu_1 =$ the true mean number of words people with normal hearing can correctly repeat from list 1, $\mu_2 =$ the true mean number of words people with normal hearing can correctly repeat from list 2, $\mu_3 =$ the true mean number of words people with normal hearing can correctly repeat from list 3, and $\mu_4 =$ the true mean number of words people with normal hearing can correctly repeat from list 4. *Plan:* If conditions are met, we should perform an ANOVA *F* test to compare the means. Random: The subjects were randomly assigned to the lists. 10%: Not necessary because there was no sampling without replacement. Normal/Large Sample: We are told that this condition is met. Same SD: The sample standard deviations for the four lists are $s_1 = 7.409$, $s_2 = 8.058$, $s_3 = 8.316$, and $s_4 = 7.779$. These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{8.316}{7.409} = 1.122 < 2.$ The conditions for ANOVA are met. *Do:* From the Minitab output we have $F = 4.92$ and *P*-value = 0.003. *Conclude:* Because the *P*-value of 0.003 is less than α = 0.05, we reject the null hypothesis. We have convincing evidence that the true mean number of words that people with normal hearing can repeat correctly is different for at least one of the 4 lists.

13.22 *State:* We want to perform a test of $H_0: \mu_1 = \mu_2 = \mu_3$ versus H_a: not all of μ_1, μ_2, and μ_3 are equal, where $\mu_1 =$ the true mean lightness measure of cloth dyed using method A, $\mu_2 =$ the true mean lightness measure of cloth dyed using method B, and $\mu_3 =$ the true mean lightness measure of cloth dyed using method C. *Plan:* If conditions are met, we should perform an ANOVA *F* test to compare the means. Random: The pieces of cloth were randomly assigned to the dyeing method. 10%: Not necessary because we are not sampling without replacement. Normal/Large Sample: Boxplots of the data for each group are given below. There is no strong skewness or outliers in any of the plots.



Same SD: The sample standard deviations for the three treatment groups are $s_1 = 0.392$, $s_2 = 0.255$, and $s_3 = 0.250$. These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{0.392}{0.250} = 1.568 < 2.$ The conditions for ANOVA are met. *Do:* From the Minitab output we have $F = 9.53$ and *P*-value = 0.001. *Conclude:* Because the *P*-value of 0.001 is less than α = 0.05, we reject the null hypothesis. There is convincing evidence to conclude that the true mean lightness measure is different for at least one dye method.

13.23 (a) For the cold temperature $\frac{s_C}{\sqrt{n}} = 8.08$ and $n = 16$ so $s_C = 8.08\left(\sqrt{16}\right) = 8.08(4) = 32.32.$ Similarly, for the neutral temperature $s_N = 5.61\left(\sqrt{38}\right) = 5.61(6.16) = 34.56,$ and for the hot temperature $s_H = 4.10\left(\sqrt{75}\right) = 4.10(8.66) = 35.51.$ These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{35.51}{32.32} = 1.099 < 2.$

(b) First we calculate the overall mean $\bar{x} = \frac{16(28.89) + 38(32.93) + 75(32.27)}{16 + 38 + 75} = \frac{4133.83}{129} = 32.05.$ Next we calculate

$$MSG = \frac{16(28.89 - 32.05)^2 + 38(32.93 - 32.05)^2 + 75(32.27 - 32.05)^2}{3 - 1} = \frac{192.8268}{2} = 96.4134.$$

Finally we compute $MSE = \frac{15(32.32)^2 + 37(34.56)^2 + 74(35.51)^2}{129 - 3} = \frac{153{,}172.35}{126} = 1215.65.$ This leads us to the statistic $F = \frac{96.4134}{1215.65} = 0.079.$ Using an *F* distribution with 3 – 1 = 2 numerator degrees of freedom and 129 – 3 = 126 denominator degrees of freedom, we find that the *P*-value is 0.924. Because this *P*-value is greater than α = 0.05, we fail to reject the null hypothesis. We don't have convincing evidence that at least one of the true mean weights at hatching for pythons hatched under these three conditions is different.

13.24 (a) For the cold temperature $\frac{s_C}{\sqrt{n}} = 5.67$ and $n = 16$ so $s_C = 5.67\left(\sqrt{16}\right) = 5.67(4) = 22.68.$ Similarly, for the neutral temperature $s_N = 4.24\left(\sqrt{38}\right) = 4.24(6.16) = 26.12,$ and for the hot temperature $s_H = 2.70\left(\sqrt{75}\right) = 2.70(8.66) = 23.38.$ These standard deviations satisfy our rule of thumb $\frac{\text{largest } s}{\text{smallest } s} = \frac{26.12}{22.68} = 1.152 < 2.$

(b) First we calculate the overall mean $\bar{x} = \frac{16(6.40) + 38(5.82) + 75(4.30)}{16 + 38 + 75} = \frac{646.06}{129} = 5.01.$ Next we calculate $MSG = \frac{16(6.40 - 5.01)^2 + 38(5.82 - 5.01)^2 + 75(4.30 - 5.01)^2}{3 - 1} = \frac{93.6529}{2} = 46.826.$

Finally we compute $MSE = \frac{15(22.68)^2 + 37(26.12)^2 + 74(23.38)^2}{129 - 3} = \frac{73{,}409.35}{126} = 582.61.$ This leads us to the statistic $F = \frac{46.826}{582.61} = 0.08.$ Using an *F* distribution with 3 – 1 = 2 numerator degrees of freedom and 129 – 3 = 126 denominator degrees of freedom, we find that the *P*-value is 0.923. Because this *P*-value is greater than α = 0.05, we fail to reject the null hypothesis. We do not have convincing evidence that at least one true mean propensity to strike for pythons hatched under these three conditions is different.

13.25 (a) The two-sample *t* statistic is $t = \dfrac{2.31 - 2.37}{\sqrt{\dfrac{(0.38)^2}{13} + \dfrac{(0.58)^2}{17}}} = -0.341$. Using technology and 27.47 degrees of freedom, the *P*-value is 0.7354. The conservative method for finding the *P*-value is to use the smaller sample size minus 1 as the degrees of freedom. Here that means using $13 - 1 = 12$ degrees of freedom. Using Table B, this gives us a two-sided *P*-value of greater than 2(0.25) = 0.50.

(b) First we calculate the overall mean $\bar{x} = \dfrac{13(2.31) + 17(2.37)}{13 + 17} = \dfrac{70.32}{30} = 2.344$. Next we calculate $MSG = \dfrac{13(2.31 - 2.344)^2 + 17(2.37 - 2.344)^2}{2 - 1} = \dfrac{0.02652}{1} = 0.0265$. Finally we compute $MSE = \dfrac{12(0.38)^2 + 16(0.58)^2}{30 - 2} = \dfrac{7.1152}{28} = 0.2541$. This leads us to the statistic $F = \dfrac{0.0265}{0.2541} = 0.104$. Using an *F* distribution with 2 – 1 = 1 numerator degrees of freedom and 30 – 2 = 28 denominator degrees of freedom, we find that the *P*-value is 0.749.

(c) The two *P*-values are quite close. We found a *P*-value of 0.739 using the two-sample *t* test and a *P*-value of 0.749 using the ANOVA *F* test. If we had used the pooled two-sample *t* test, we would have found that the *P*-values were identical.

13.26 c

13.27 c

13.28 b

13.29 d

13.30 Answers will vary.